Military Operations Research Society

73rd MORSS CD Cover Page

712CD

For office use only 41205

UNCLASSIFIED DISCLOSURE FORM CD Presentation

21-23 June 2005, at US Military Academy, West Point, NY

Please complete this form 712CD as your cover page to your electronic briefing submission to the MORSS CD. Do not fax to the MORS office.

Author Request (To be completed by applicant) - The following author(s) request authority to disclose the following presentation in the MORSS Final Report, for inclusion on the MORSS CD and/or posting on the MORS web site.

Name of Principal Author and all other author(s): Thomas F. Chwastyk

Principal Author's Organization and address:

AFSAA/SAAT
1777 N Kent St
Rosslyn VA 22209

Phone: DSN 425-8672 (Com 703-588-8672)
Fax: DSN 425-8738 (Com 703-588-8738)
Email: Thomas.Chwastyk@pentagon.af.mil

Original title on 712 B: Design of Experiments (DOE) – Its Role in Validation of STORM Based on THUNDER

Revised title: (No change)

Presented in (input and Bold one): **WG 25 & 29**, (CG, Special Session, Poster, Demo, Tutorial): N/A

This presentation is believed to be:
UNCLASSIFIED AND APPROVED FOR PUBLIC RELEASE

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
22 JUN 2005	**N/A**	-

4. TITLE AND SUBTITLE
Design of Experiments (DOE) for Validation of STORM Based on THUNDER

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFSAA/SAAT 1777 N Kent St Rosslyn VA 22209

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM201946, Military Operations Research Society Symposium (73rd) Held in West Point, NY on 21-23 June 2005. , The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**41**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Headquarters U.S. Air Force

Integrity - Service - Excellence

Design of Experiments (DOE) for Validation of STORM Based on THUNDER

73rd MORSS, USMA 2005
WG 25 & 29

Mr. Tom Chwastyk
AFSAA/SAAT

U.S. AIR FORCE

STORM

Purpose

- THUNDER vs. STORM
- Goal of DOE here: V&V by comprehensive comparison
  - Why not VV&A?
- General and specific models of the V&V experiment
- DOE for adequate excitation to compare responses
  - Screening to reduce response model size
  - Adding design points to check for interactions
- Comparing THUNDER and STORM response
  - By Singular Value Decomposition
  - By Canonical Correlation
- Automating execution of runs and extraction of results
  - Work in progress – all feedback welcome
- Punch line: Massive quantitative validation experiment – uses computers and sophisticated techniques to exceed past human limits on data production and interpretation

THUNDER vs. STORM

- THUNDER: accepted standard
  - Not Government owned
  - Proprietary engine, SIMSCRIPT
    - CACI licenses – big $$$ / seat / year
  - Ground war:
    - "Pistons" define FLOT
    - Piston walls do not reorient as FLOT distorts
  - No study manager -- manual file configuration control
  - Minimal GUI
  - Minimal ISR

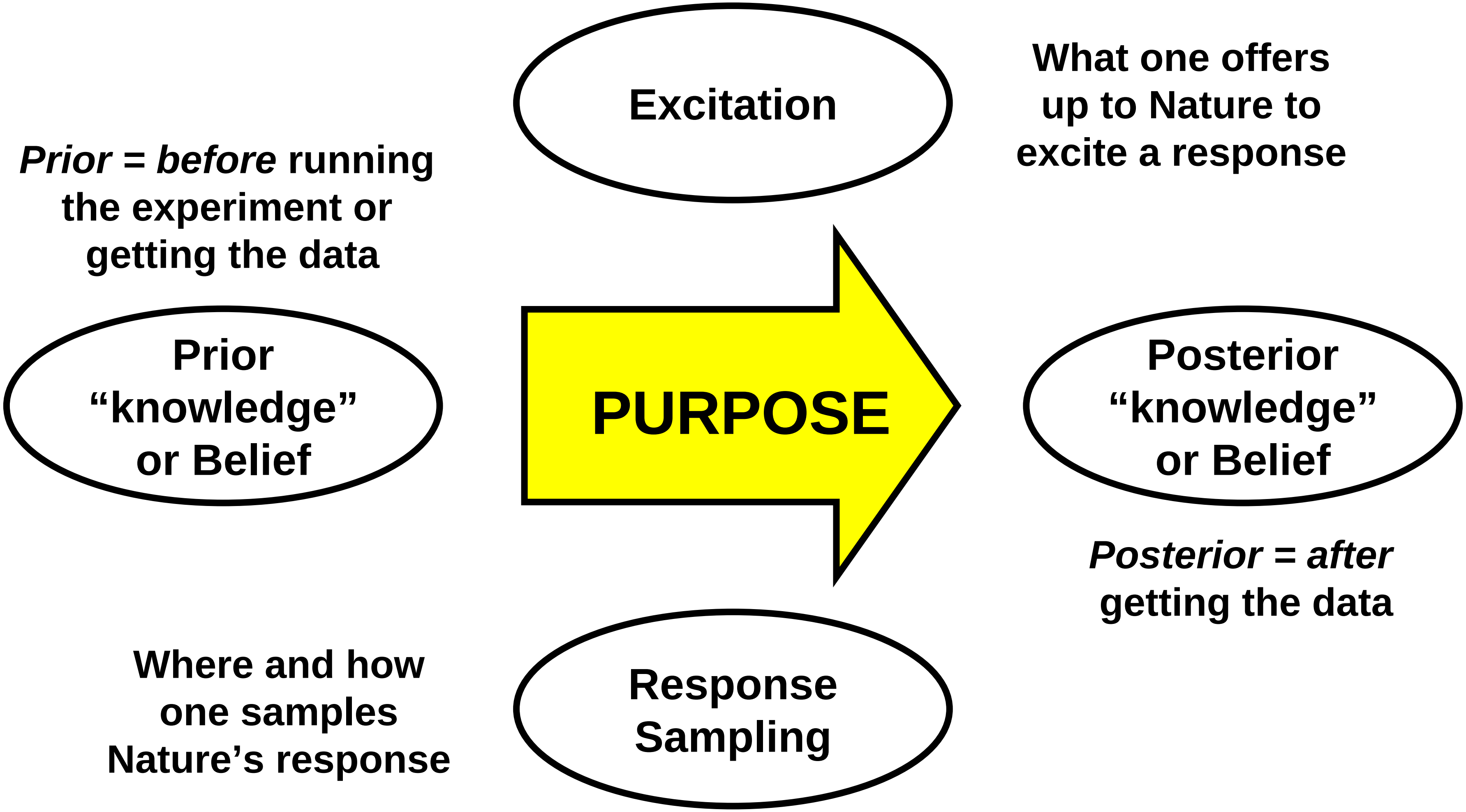
- STORM: AFSAA's next standard
  - Government owned
  - Open source, non-proprietary
    - Simulation engine in g++ ; maps, DB tools; even OS, all open
    - No per-seat license fees
  - Ground war:
    - Along arcs and within nodes
    - Shape of FLOT morphs naturally under network constraints
  - Study manager automates file configuration control
  - Extensive GUI – map tool, report tool, graph tool
  - Sat tool and other significant ISR developments

U.S. AIR FORCE

V&V by Comparison (Not VV&A?)

- V&V
  - Verification: "Solve the Problem Right"
    - Identify and fix lurking code problems
  - Validation: "Solve the Right Problem"
    - Accepted referent is THUNDER but comparison is a hard problem
  - Comparison needed <u>over a set of scenarios of interest</u>
    - Are results "comparable to or better than" THUNDER? Just what does that mean?
      - SME's "face validation" vs. statistical tools – both are needed
    - Does DOE show effects of changes in STORM vs. THUNDER?
- Why V&V, not VV&A?
  - A (Accreditation) is PM's decision – balancing risk versus resources
    - Too little V&V: low resource expenditure but high risk
    - Too much V&V: little risk reduction for last expenditures

A General Model of an Experiment



…Tailored to V&V and Comparison

STORM uses next generation THUNDER campaign model

Excitation

Excitation is the set of scenarios, each with its experimental design and with levels for factors: <u>the INPUTS</u>

Prior knowledge or Belief

PURPOSE

Posterior knowledge or Belief

Response Sampling is MOMs + dofs extracted from simulation: <u>the OUTPUTS</u>

Response Sampling

STORM improves on THUNDER – it's a good replacement

V&V Scenario Set of Interest

- Consult with SMEs, see for what campaign models/scenarios THUNDER results (inputs and outputs) are already available and of interest
  - THUNDER has been much more expensive to run (small error $\Rightarrow$ typically 60-100 reps) than STORM – hence, minimize additional THUNDER runs
- Choose among available THUNDER campaigns to agree on sufficiently complete V&V scenario set
- If necessary, add THUNDER or STORM work to achieve common and complete set of scenarios of significant interest for V&V

U.S. AIR FORCE

Need to Reduce Input Dimensions

THUNDER	v6.9	99	*.dat files
STORM	v1.3	115	*.dat files (plus path.dat)
- Typical .dat file has 10 to 1000+ parameters (inputs) – highly interrelated, often nonlinearly
- Too many parameters to screen individually
  - Use SME's to combine (cull?) inputs into fewer (10 to 100) factors $f_i \equiv_{def} [\ p_{i,j'}\ ]$ parameter vectors
  - $p_{i,j}$, $p_{i,k\neq j}$ need not be in same .dat file
  - Each parameter in a factor has two levels, A / B – goes A / B as its factor goes high / low – no "within-factor interactions"
- Factor screening takes place in STORM across entire V&V scenario set at once
- Whether the change in level of factor f_i is significant depends both on the scenario set and on what MOM(s) or output dof(s) are of interest

U.S. AIR FORCE

Input Sensitivity Screening

- **Initial screening in STORM of two-level factors**
  - **Resolution V (preferred) or IV (if necessary) fractional factorial (FF) – possibly even res. III for initial stab?**
    - **If resolution IV necessary, follow up with resolution V (res. IV significant factors only, plus all 2-factor interactions) to screen for significant interactions**
  - **256 reps (resolution IV) $\Rightarrow$ up to 128 factors**
    - **Comparable to # of reps in STORM V&V reports for previous releases (19 + cases x 10 repeated reps / case)**
    - **Likely fewer factors ($\leq$64, $\leq$32) if long reps (t_{sim}>>10 days)**
  - ***However*, 256 reps (resolution V) $\Rightarrow$ only 12 to 17 factors (+ all 2-level interactions)**
- **In any case –**
  - **NOT One-Factor-At-a-Time as before!**
  - **NO replicated design points (reps)!**

U.S. AIR FORCE

Matrix Model Equations

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$$

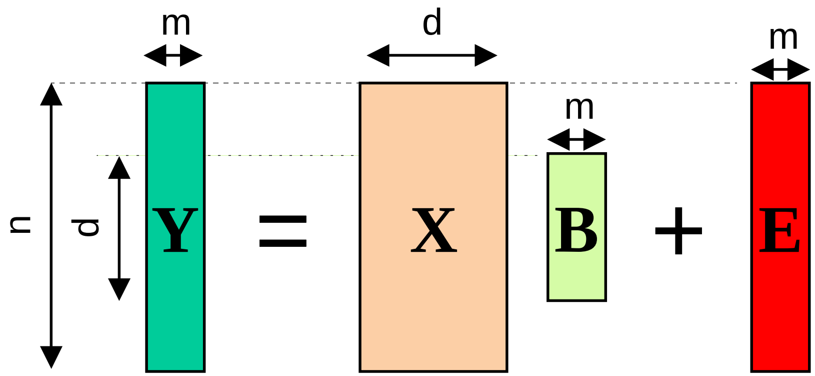
$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} + \hat{\mathbf{E}}(\mathbf{s})$$

$$\hat{\mathbf{E}}'_i = \left[f_1(s_i) \ \cdots \ f_j(s_i) \ \cdots \ f_m(s_i) \right]'$$

$\hat{\mathbf{B}}$ is by least squares, minimizes

$$rss = \sum_i \sum_j (y_{ij} - \hat{y}_{ij})^2$$

Not Your Usual Few Responses



m = # of response variables

d = # of design variables (factors)

n = # of runs

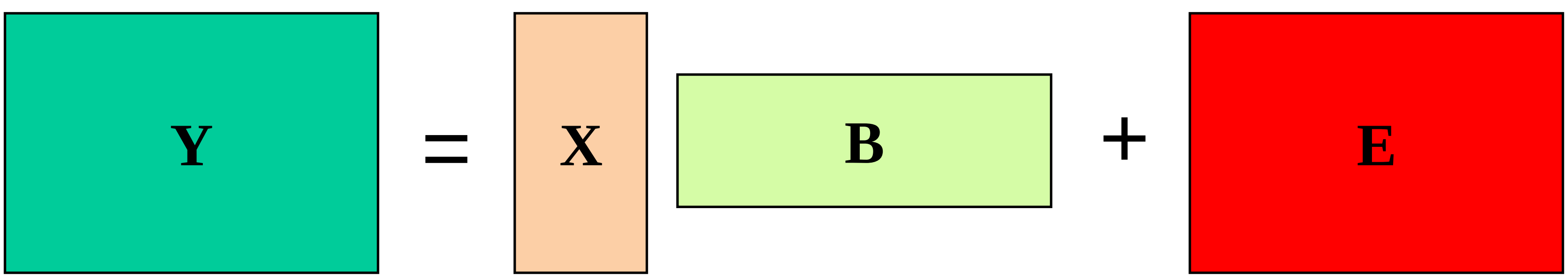
n > d (always)

m << d (usually)

m >> d

Why does Y need to be so broad, m >> d ?

In our V&V experiment, X is <u>excitation</u>; Y samples <u>response</u>. We are deliberately oversampling.



We oversample by several times the dimensionality of our excitation in a “shotgun approach”, hoping we catch most of the significant dimensions of model response.

U.S. AIR FORCE

Why Fit Is Always Too Good

$$\mathbf{Y} = \mathbf{X}\,\mathbf{B} + \mathbf{E}$$

$$\hat{\mathbf{Y}} = \mathbf{X}\,\hat{\mathbf{B}} + \hat{\mathbf{E}}$$

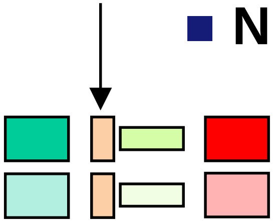
$$\mathbf{X}'\,\mathbf{E} \neq 0 = \mathbf{X}'\,\hat{\mathbf{E}}$$

Elements of Matrix Model

- Starts as typical multivariate response model
  - Response matrix $\mathbf{Y}$ (n×m)
  - Design matrix $\mathbf{X}$ (n×d)
  - Effects matrix $\mathbf{B}$ (d×m)
  - "Out-of-model" matrix $\mathbf{E}$ (n×m)
  - Estimates "hatted": $\hat{\mathbf{E}}$, $\hat{\mathbf{B}}$
- Beyond typical model:
  - *Compute* $\mathbf{Y}$ from $\mathbf{X}$ in permuted order $pco^{(1)}$, $pco^{(2)}$, $pco^{(3)}$, but *record* $\mathbf{Y}$ in design order (picture of $\mathbf{Y}$, pco_i, $pco^{(i)}$ later)
  - Record random seed and permuted computational order $[\, s_i \;\; pco_i \,]$ for row i (nx2)

Design Matrix X - Words

- ROWS in the design matrix X
  - Give particular combination of factor levels for each design point
  - Cover all scenarios of interest
  - Contain a non-zero entry in only one of the columns indicating scenarios
- COLUMNS in the design matrix X
  - Signify design (input) factors
  - Signify scenarios for estimating mean
    - Decompose sum of squares due to scenario means
    - Number of means is included in d, # of design variables



Design Matrix - Picture

X

scenario # 1	2	3	4	5	6	7	8	9	10 factor level
1	0	0	0	1	3	5	7	2	8
1	0	0	0	2	4	6	7	2	8
1	0	0	0	2	4	5	7	2	8
1	0	0	0	1	3	6	7	2	8
0	1	0	0	1	3	5	8	7	1
0	1	0	0	2	4	6	8	7	1
0	1	0	0	2	4	5	8	7	1
0	1	0	0	1	3	6	8	7	1
0	0	1	0	1	3	6	8	7	8
0	0	1	0	1	3	5	8	2	1
0	0	1	0	1	3	6	7	7	8
0	0	1	0	1	4	5	7	2	1
0	0	1	0	1	4	6	8	7	8
0	0	1	0	1	4	5	7	2	1
0	0	0	1	1	4	6	8	2	8
0	0	0	1	2	3	6	7	2	8
0	0	0	1	1	3	6	8	2	1
0	0	0	1	2	3	5	7	7	1
0	0	0	1	1	3	5	7	7	8
0	0	0	1	2	4	5	8	7	8

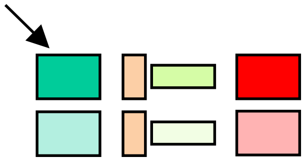
1's account for scenario means, ssq

constant factor level within scenario says significance of factor is not tested within scenario

Notional design only, not 2 $^{k-p}$

Response Matrix Y - Words

- ROWS in the response matrix Y
  - Correspond to design matrix rows
- COLUMNS in the response matrix Y signify
  - Scalar or vector MOMs / dofs
  - Consistent responses for each row and scenario
    - Squadrons or other entity types handled differently by planning files on different design points should not re-use the same response columns unless
      - the difference in response due to difference in planning file parameters is the effect being sought and
      - the difference in planning file parameters is reflected in one or more factors



U.S. AIR FORCE

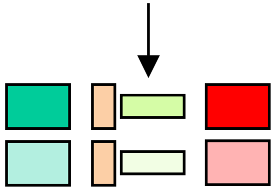
Response Matrix - Picture

$pco^{(i)}$	pco_i	scenario #	i = row # in X	Scalar MOM 1	Scalar MOM 2	Scalar MOM 3	Vector dof 1 day 1	Vector dof 1 day 2	Vector dof 1 day 3	…	Vector dof 1 day n-1	Vector dof 1 day n	Scalar MOM 4	Vector dof 2 day 1	Vector dof 2 day 2	Vector dof 2 day 3	…	Vector dof 2 day n-1	Vector dof 2 day n	Scalar MOM 5	Vector dof 3 day 1	Vector dof 3 day 2	…	Vector dof 3 day n-1	Vector dof 3 day n	Vector dof 4 day 1	… etc.
13	14	1	1																								
11	3	1	2																								
2	5	1	3																								
7	9	1	4																								
3	11	2	5																								
10	12	2	6																								
12	4	2	7																								
8	8	2	8																								
4	13	3	9																								
14	6	3	10																								
5	2	3	11																								
6	7	3	12																								
9	1	3	13																								
1	10	4	14																								
…	…	…	…																								

$\mathbf{Y}$

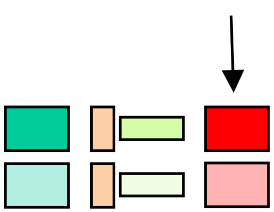
Effects Matrix B, B hat

- ROWS in the effects matrix B
  - Correspond to design variables
  - For SME-aggregated factors, coefficient shows effect of change in level
  - For scenario indicator variables, coefficient shows scenario mean for each response
- COLUMNS in the effects matrix B
  - Correspond to response MOMs, dofs



Out-of-Model Matrix E, E hat

- **Part of E is assumed due to stochastic effects**
  - **Neither random seed value s_i nor its effect $f_j(s_i)$ on each response y_{ij} is known to the fitting process**
  - **s_i and $f_j(\bullet)$ — hence $f_j(s_i)$ — are termed exogenous (excluded from model)**
- **Effects of exogenous variables, *including stochastic effects*, alias to an unknown extent onto endogenous (included in model) effects b_j**



U.S. AIR FORCE

Interactions, Response Surface

- Once factors and 2-factor interactions are identified as significant at resolution V, high-low level only …
- Add intermediate levels as appropriate to gain Response Surface Model (RSM)
  - Central composite design if appropriate
- Retain only <u>statistically</u> significant factors (and interactions) for comparing STORM to THUNDER
  - Not normal statistical practice to "cherry pick" interactions without physical model justification
  - Empirically justifiable because RSM is <u>descriptive</u> only, not first principles
  - We are a long way off from this

U.S. AIR FORCE

Dimension Reduction (1): RSM

- **Add face and center points for significant factors to get central composite design if feasible**
- **Possibly use intermediate points for near-orthonormal Latin Square**
- **Get response surface model (RSM) – polynomial or spline**
  - **Stop on Mallow's C_p or other measure of complexity vs. goodness of fit**
- **<u>Retains coordinates of original model space</u>**

U.S. AIR FORCE

Weight Questions (1)

- **Question 1: What is the proper weight or scale per MOM / dof(s) in the penalty function for lack of fit?**
  - **That is, how much should lack of fit in one column (or set of columns) of Y count relative to lack of fit in another column (or set of columns)?**
- **Question 2: Same as Question 1 but for one scenario (set of rows) versus another**

U.S. AIR FORCE

Weight Answers (Partial) (2)

- Depends on user goal and on STORM / THUNDER comparison approach
  - Data for SVD, canonical correlation should have zero scenario mean
  - Appears no natural scale for Singular Value Decomposition (SVD); use unit column variance for consistency with canonical correlation
  - Scale set to unit column variance by definition of canonical correlation, but meaning and appropriateness to user goal are not clear
  - Beware augury / mysticism / wise chin-scratching

U.S. AIR FORCE

Equations in Passing

- Final dry run for this brief commented that there were way too many equations, too few graphics
- Graphics proved far more time consuming to produce than initially expected
- Equations were retained more as art objects and speaking totems than as detailed communication tools
- Equations past the basic “Matrix Model Equations” will not be on the midterm or the final

U.S. AIR FORCE

Dim. Red. (2): SVD (1)

- **Singular Value Decomposition (SVD) on responses Y adjusted to zero-scenario-mean, scaled to unit Y column norm, and transposed*:**

* not a good idea, but not fatal

$$\underset{(n\times m)}{\mathbf{Y}_{zsm}} = \mathbf{Y}(\mathbf{I}_n - \underbrace{\mathbf{X}_{sm}(\mathbf{X}_{sm}'\mathbf{X}_{sm})^{-1}\mathbf{X}_{sm}'}_{\mathbf{X}_{sm}=\text{only columns for scenario means}})$$

$$\underset{(m\times n)}{\mathbf{A}} = \left(\hat{\mathbf{Y}}_{zsm}\right)', \left(\hat{y}_{zsm}\right)_{ij} = \left(y_{zsm}\right)_{ij} \Big/ \sqrt{\sum_{i'=1}^{n}\left(y_{zsm}\right)_{i'j}^2}$$

$$\mathbf{A} = \mathbf{ODV}' \begin{cases} \mathbf{O}'\mathbf{O} = \mathbf{I}_n, (m > n) \Rightarrow \mathbf{OO}' \neq \mathbf{I}_m \\ \mathbf{D} \text{ diagonal, p.s.d.}, (i < j) \Rightarrow (d_{ii} \geq d_{jj}) \\ \mathbf{VV}' = \mathbf{V}'\mathbf{V} = \mathbf{I}_n \end{cases}$$

… SVD (2)

- **Continuing,**

$$\mathbf{A}'\mathbf{A} = \mathbf{Z} = \mathbf{V}\mathbf{D}'\mathbf{O}'\mathbf{O}\mathbf{D}\mathbf{V}'$$
$$= \mathbf{V}\mathbf{D}^2\mathbf{V}'$$
$$\mathbf{Z}\mathbf{V} = \mathbf{V}\mathbf{D}^2\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{D}^2$$
$$\mathbf{Z}\mathbf{V} = \mathbf{V}\boldsymbol{\Lambda} \quad \text{(symmetric eigenproblem)}$$
$$d_{ii} = \sqrt{\lambda_i} \quad (\text{sorted } \lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0)$$
$$\mathbf{O}_i = \mathbf{A}\mathbf{V}_i d_{ii}^{-1} \quad \text{canonical column } i \text{ of } \mathbf{A} \text{ for } d_{ii} > 0$$

- **Why is this significant?**

… SVD (3)

- This is significant *because*…

$$\mathbf{A} = \mathbf{ODV'} = \sum d_{ii} \mathbf{OV'} \text{ (}\underline{\text{exactly!}}\text{) and}$$

$$\sum_j \sum_k (\mathbf{O}_i \mathbf{V}_i')^2_{jk} = \sum_j \sum_k (o_{ji} v_{ik})^2 = \sum_j o^2_{ji} \sum_k v^2_{ik} = 1,$$

so <u>all scale information is in the "singular values"</u> d_{ii} in the sense that d_{ii}^2 shows how many unit rows of A are expressed by $d_{ii}\mathbf{V}'_i$.

- This means that the completeness $R^2(k,m)$ of expressing all m unit rows of the response matrix A using only k coefficients is

$$R^2_{k,m} = \left(\sum_{i=1}^{k} d^2_{ii}\right) \Big/ \left(\sum_{i'=1}^{m} d^2_{i'i'}\right)$$

- More on SVD follows after discussion on Canonical Correlation

DR (3): Canonical Correlation (1)

- Two response sets, say $\mathrm{T}=\hat{\mathrm{Y}}'_{\mathrm{T}}$ for THUNDER and $\mathrm{S}=\hat{\mathrm{Y}}'_{\mathrm{S}}$ for STORM
- *m* ROWS of T, S
  - Zero scenario mean, unit variance as for SVD
  - Same row ⇔ same MOM or dof for T as for S
- *n* COLUMNS of T, S are in same order as design matrix rows, although design points were computed in permuted sequence(s)

U.S. AIR FORCE

… CanCorr (2)

- **Canonical correlation (CanCorr) gives** $m\times 2$ **coefficients** $[\,\mathbf{lca}_k\ \mathbf{lcb}_k\,]$ **for successive** $n\times 2$ **linear composites** $[\,\mathbf{u}_k\ \mathbf{w}_k\,]$**, such that for** $\mathbf{u}_k' = \mathbf{lca}_k'\mathbf{T}$, $\mathbf{w}_k' = \mathbf{lcb}_k'\mathbf{S}$, $k = 1, 2, \ldots, \min(m,n)$
  - $\mathbf{u}_k$ **and** $\mathbf{w}_k$ **have maximal correlation** ρ^*_k **between any two linear composites of** T, S **<u>not</u> spanned by** $[\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_{k\text{-}1}]'$, $[\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_{k\text{-}1}]'$
  - $\mathbf{Var}(\mathbf{u}_k) = \mathbf{Var}(\mathbf{u}_k) = 1$
  - $\rho^*_1 \geq \rho^*_2 \geq \rho^*_3 \geq \ldots \rho^*_{\min(m,n)} \geq 0$
  - **For** $j > k$, $\mathbf{u}_j$ **and** $\mathbf{w}_k$ **are uncorrelated, i.e.,**

$$\mathbf{u}_j'(\mathbf{I}-\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}')\,\mathbf{w}_k = 0$$

U.S. AIR FORCE

… CanCorr (3)

- How does this work? From the zero scenario means, unit row norms, and definition of variance we get

$$\underset{(2m\times 1)}{\mathbf{0}} = \begin{bmatrix} \mathbf{T} \\ \mathbf{S} \end{bmatrix} \underset{(n\times 1)}{\mathbf{1}}$$

$$\underset{(2m\times 2m)}{\mathbf{\Sigma}} = \frac{1}{n - n_{zsm}} \begin{bmatrix} \mathbf{T} \\ \mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{T} \\ \mathbf{S} \end{bmatrix}'$$

$$= \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{TT}} & \mathbf{\Sigma}_{\mathbf{TS}} \\ \mathbf{\Sigma}_{\mathbf{ST}} & \mathbf{\Sigma}_{\mathbf{SS}} \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$$

U.S. AIR FORCE

... CanCorr (4)

- Then it can be shown in about 3 textbook pages that

$$\begin{bmatrix} \mathbf{u}_k \\ \mathbf{w}_k \end{bmatrix} = \begin{bmatrix} \underbrace{\mathbf{e}_k{}'\mathbf{\Sigma}_{11}^{-1/2}}_{\mathbf{lca}_k}\mathbf{T} \\ \underbrace{\mathbf{f}_k{}'\mathbf{\Sigma}_{22}^{-1/2}}_{\mathbf{lcb}_k}\mathbf{S} \end{bmatrix}, \text{ where}$$

$$\text{Corr}(\mathbf{u}_k, \mathbf{w}_k) = \rho_k^* \text{ is the } k^{\text{th}} \text{ largest eigenvalue of}$$

$$\boxed{\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2}}, \text{ with eigenvector } \mathbf{e}_k, \text{ and also of}$$

$$\boxed{\mathbf{\Sigma}_{22}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2}}, \text{ with eigenvector } \mathbf{f}_k.$$

U.S. AIR FORCE

… CanCorr (5)

- First approach will be to look at initial canonical variables (cv's) as long as ρ_k are well separated
  - If ρ_k are closely spaced, eigenvectors and hence cv's are determined only up to rotations
- Second approach will be to look at residuals once projections on all cv's from 1 to k are removed, for k's such that
  - k is not in the middle of a closely spaced block of ρ_k
  - k is not so far beyond 1 that any patterns apparent in residuals are as likely to be artifacts of numerical methods as of remaining THUNDER vs. STORM relationships
- **We have not done this before; any advice would be welcome.**

SVD versus CanCorr: SVD (4)

- **A'A = Z of the SVD development can now be seen as Σ_{22} of CanCorr.**
- **We also have Σ_{12}, the covariation of T with S. We can estimate what we would expect for T based on S using the first k singular vectors of S as**

$$\hat{\mathbf{T}}(\mathbf{S},k) = \mathbf{\Sigma}_{12}\left(\mathbf{\Sigma}_{22}\right)^{-1|k} = \mathbf{\Sigma}_{12}\mathbf{O}\begin{pmatrix} d_{11}^{-1} & 0 & \cdots & & & \\ 0 & \ddots & & & & \\ \vdots & & d_{kk}^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}\mathbf{O}'$$

SVD versus CanCorr: SVD (5)

- Similar remarks and approaches as for CanCorr apply to SVD
  - Singular vectors are less well defined as eigenvalues become closely spaced
  - We may look at the fits (estimated values) first if most of the significance falls into a few numerically stable patterns
  - We will look at residuals as well, particularly at breaks in a block-type pattern of decrease in singular values / eigenvalues

SVD versus CanCorr: SVD (6)

- **The big, big difference between SVD and CanCorr is**
  - **SVD aims to compactly express the variation of a *single* data set**
  - **Covariation with the other data set is dealt with almost as an afterthought.**
- **CanCorr aims to compactly express the covariation of *both* data sets, without favor for one or the other.**

Automating Runs and Results (1)

- Again, a work in progress – all feedback welcome
- Number of parameters, factors, runs makes manual DOE setup and execution unattractive for all but small problems
  - Manual "baby steps" on small problems for debugging
- Concept is to factor problem into DOE files giving
  - Design matrix, source citation, permuted computational order
  - Factors and associated parameters, with .DAT file levels for parameters as .DOE factor goes between its levels

U.S. AIR FORCE

Automating Runs and Results (2)

- Automate run process
  - Substituting parameter .DAT levels into .DOE files
  - Write result to temporary .DAT files
  - Execute runs in permuted order
- Automate DOE post process
  - Extract results
  - Transfer to statistical software
  - Use stat tools for evaluation and comparison between THUNDER, STORM results on nominally equivalent models

U.S. AIR FORCE

Punch Line: Quantitative V&V?

- This is a massive quantitative validation experiment
  - Uses computers to exceed past human limits on data production
  - Uses computers and sophisticated math to surpass unaided human limits on data interpretation
  - Unexpected differences between THUNDER and STORM will constitute new knowledge
  - Finding only differences explainable as due to known deliberate changes would be a disappointment

Nothing to Lose But Uncertainty

- The author's involvement stems from > 20 years trying to compare test data with modeling results
  - 16 years shock testing nuclear powered ships and their components
  - 9 years with robotic test equipment testing composite material systems
    - Assumed material response theory (dissipated energy density) said could account for 60K data points via 125 coefficients and constraints
    - Ability to so compactly model test data gave massive support to underlying theory
- It is long past time for a similar quantitative approach to V&V – and comparing THUNDER and STORM should be a milestone in that effort.

Questions?